Econometric Analysis

Ryuichi Tanaka

Tokyo Institute of Technology

What is the Econometrics?

- An empirical analysis uses data to test a theory (not necessarily economics, but sociology, history, biology etc.) or to estimate a relationship (any relationship)
- A formal (economic) model can be tested
- Theory may be ambiguous as to the effect of some policy change – can use econometrics to evaluate the program (policy change)

Example 1: Human Capital Theory

- Human capital theory (Becker): Education increases the skills and thus earnings of the individual
  - Education: investment in human capital
  - Labor earnings: the returns from the investment
- Relationship: Educational attainment (years of education) and the annual labor income

Example 2: R&D and value of firms

- R&D investment theory: R&D investment increases the future profitability of the firm thus the value of the firm
- Relationship: R&D investment flow and the stock price of the firm
- Question: Does R&D investment increase the stock price ?

Example 3: Police and city crime rate

- Theory: An increase in the number of police per capita mitigates crime rate
- Relationship: the number of police per capita and murder rate (the number of murders per capita) in a city

Other Examples

- The degree of income inequality and economic growth rate (Does lower inequality lead to faster economic growth?)
- Job training program and labor income (does the participation to a job training program increase labor income?)

Data Type１：Cross-sectional data

- Cross-sectional data is a random sample
- Each observation is a new individual, firm, country etc. with information at a point in time
- If the data is not a random sample, we have a sample-selection problem

Table 1.1: A Cross-Sectional Data Set on Wages and Other Individual Characteristics

obsno	*wage*	*educ*	*exper*	*female*	*married*
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Table 1.2: A Data Set on Economic Growth Rates and Country Characteristics

obsno	country	gpcrgdp	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
61	Zimbabwe	2.30	17	6

Data Type 2 : Time-series data

- Time series data has a separate observation for each time period – e.g. stock prices
  - NIKKEI 225 from January 1st, 1990 to December 31st, 2008
  - GDP and government expenditure of Japan from 1945 to 2008
  - Height and weight of you from the birthday to today
- Since time-series data is not a random sample, we need to consider different problems
- In particular, trends and seasonality will be important

Table 1.3: Minimum Wage, Unemployment, and Related Data for Puerto Rico

obsno	year	avgmin	avgcov	unemp	gnp
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

Data Type 2 : Panel data

- Hybrid of cross-sectional and time-series data
  - NIKKEI 225 and Dow Jones Industrial Average from January 1st, 1990 to December 31st, 2008
  - GDP and government expenditure of OECD countries from 1945 to 2008
  - Height and weight of all of you from the birthday to today
- Can follow the same random individual observations over time – known as panel data or longitudinal data

Table 1.5: A Two-year Panel Data Set on City Crime Statistics

obsno	city	year	murders	population	unem	police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
297	149	1986	10	260700	9.6	286
298	149	1990	6	245000	9.8	334
299	150	1986	25	543000	4.3	520
300	150	1990	32	546200	5.2	493

The Question of Causality

- The ultimate goal of empirical study: What is the causal relationship between the two events of interest?
  - Determine whether a change in one variable causes a change in another variable
  - Simply establishing a relationship between variables (correlation) is rarely sufficient
- This is imperative for policy evaluation

Example (of simultaneity): Police and city crime rate

- Population in Washington D.C. (588,292) is same as in Denver, Colorado (588,349) in 2007
- Number of police officers in Washington D.C. (4400) is a triple of Denver (1548) in 2007
- Number of homicide in Washington D.C. (181) is a triple of Denver (50) in 2007
- Correlation between the number of police officers and the number of homicide is positive
- Policy implication: reduce the number of police officers in order to mitigate homicide!?

Example (of omitted variables): the returns to education

- Question: Does educational raise labor income?
  - Answer from the theory: YES
- The simple regression model: $Earnings = \beta_0 + \beta_1 education + u$
- The estimate of β_1, is the return to education, but can it be considered causal?
  - If the error term, u, includes other factors affecting earnings (such as IQ, unobserved ability), the positive correlation between them may be spurious
  - If the relation is spurious, scholarship provision policy is ineffective to reduce poverty (causality is necessary for policy proposal)

Causality with experimental data

- With experimental data, we can identify the causal relationship of interest (ideal benchmark)
- Experiment generates data of treatment group and control group
- Example
  - Evaluate the effect of new medicine
  - Control for age, height, weight, sex, education…
  - Treatment group: new medicine
  - Control group: placebo (no effect)
  - Naïve comparison of outcomes identifies causality

Causality with non-experimental data

- It is rare in economics (and many other areas without labs!) to have experimental data
- We need to use non-experimental, or observational, data to make inferences (all data in the examples are observational!)
- We want to control for observables (IQ, parental influences such as network, income, genetically transmitted ability, etc.) as much as possible (ceteris paribus analysis)
  - Some things are still unobserved, which can be problematic (we study how to deal with this problem in several settings)

Ceteris paribus analysis

- Ceteris paribus: holding all other relevant factors fixed
- In the above example, ceteris paribus analysis is to examine if an increase in education raises earnings, holding all other factors such as IQ, parental influences ability fixed
- Multiple regression model is a method of ceteris paribus analysis

A mathematical representation of ceteris paribus analysis: conditional expectation

- Conditional expectation: holding other control variables *c* fixed, how does a change in ***w*** affect the expectation of ***y*** ?

$$E(y \mid w, c)$$

  - How do we model this conditional expectation? How can we estimate the model?
- Partial effect: a mathematical representation of causal effect

$$\frac{\partial E(y \mid w, c)}{\partial w}$$

Reality of ceteris paribus analysis

- How can we find <u>all</u> relevant control variables? Are they all observable?
  - Almost impossible (unobserved ability is unobserved)
- It is up to you which control variables you choose, but you need to choose appropriate variables to make your analysis convincing
- (Economic) theory guides you how to choose the variables, and econometrics theory tells you if you are on the right track (this is what we learn in this course)

Basic ingredients of econometric analysis

- <u>Goal</u>: want to know the structure of the population distribution
  - Empirical goal: What is the causal relationship between the two events of interest?
  - More specific goal: want to know the conditional expectation
- <u>Approach</u>: Estimate the conditional expectation with the independently and identically distributed data drawn from the population

Theoretical concern 1: consistency

- Consistent estimator: you can see the true structure of a population model when you increase sample size indefinitely (If you can use infinite size of sample, then you should see the true structure of the population as long as you use a consistent estimator)
- Estimator: estimation method
- Consistency is a minimal requirement for an estimator
- Theoretical question: Is an estimator consistent? (Does the way of estimation give us a consistent estimator?)
- We show consistency of estimators by weak law of large number

Theoretical concern 2: asymptotic distribution

- How does the asymptotic distribution of the estimator look like?
- Using asymptotic distribution, we can carry out statistical inferences (e.g., test of significance)
- We use central limit theorem to show the asymptotic distribution is a normal or a chi-squared distribution

Problem Set 0

- Take any theory. Consider a relationship between two events (variables) of your interest. Expect the relationship between them. Discuss whether the expected relationships are causal (for at least five examples)

- Choose one example from the above. Discuss how to implement experimental study to identify the causal effect of one event on the other.